RAW SEQUENCE LISTING
PATENT APPLICATION US/09/502,664A

INPUT SET: S36742.raw

FEB 12 2002

TECH CENTER 1600/2900

RECEIVED

ENTERED

This Raw Listing contains the General
Information Section and up to the first 5 pages.

```
 1                                    SEQUENCE LISTING
 2
 3     (1)     General Information
 4
 5            (i) APPLICANT: Hadlaczky, Gyula
 6                          Szalay, Aladar
 7
 8            (ii) TITLE OF THE INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
 9     AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
10
11            (iii) NUMBER OF SEQUENCES: 34
12
13            (iv) CORRESPONDENCE ADDRESS:
14               (A) ADDRESSEE: Heller Ehrman White & McAuliffe
15               (B) STREET: 4250 Executive Square, 7th Floor
16               (C) CITY: La Jolla
17               (D) STATE: CA
18               (E) COUNTRY: USA
19               (F) ZIP: 92037
20
21            (v) COMPUTER READABLE FORM:
22               (A) MEDIUM TYPE: Diskette
23               (B) COMPUTER: IBM Compatible
24               (C) OPERATING SYSTEM: DOS
25               (D) SOFTWARE: FastSEQ Version 1.5
26
27            (vi) CURRENT APPLICATION DATA:
28               (A) APPLICATION NUMBER:
29               (B) FILING DATE: 28-NOV-2000
30
31            (vi) PRIOR APPLICATION DATA:
32               (A) APPLICATION NUMBER: 08/835,682
33               (B) FILING DATE: 10-APR-1997
34               (C) CLASSIFICATION:
35
36            (vi) PRIOR APPLICATION DATA:
37               (A) APPLICATION NUMBER:  08/695,191
38               (B) FILING DATE: 07-AUG-1996
39               (C) CLASSIFICATION:
40
41            (vi) PRIOR APPLICATION DATA:
42               (A) APPLICATION NUMBER:  08/682,080
43               (B) FILING DATE: 15-JUL-1996
44               (C) CLASSIFICATION:
45
46            (vi) PRIOR APPLICATION DATA:
```

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/502,664A

INPUT SET: S36742.raw

```
47        (A) APPLICATION NUMBER: 08/629,822
48        (B) FILING DATE: 10-APR-1996
49        (C) CLASSIFICATION:
50
51    (viii) ATTORNEY/AGENT INFORMATION:
52        (A) NAME: Seidman, Stephanie L
53        (B) REGISTRATION NUMBER: 33,779
54        (C) REFERENCE/DOCKET NUMBER: 6869-402E
55
56
57    (ix) TELECOMMUNICATION INFORMATION:
58        (A) TELEPHONE: 858-450-8403
59        (B) TELEFAX: 858-587-5360
60        (C) TELEX:
61
62        (2) INFORMATION FOR SEQ ID NO:1:
63
64    (i) SEQUENCE CHARACTERISTICS:
65        (A) LENGTH: 1293 base pairs
66        (B) TYPE: nucleic acid
67        (C) STRANDEDNESS: single
68        (D) TOPOLOGY: linear
69
70    (ii) MOLECULE TYPE: Genomic DNA
71    (iii) HYPOTHETICAL: NO
72    (iv) ANTI-SENSE: NO
73    (v) FRAGMENT TYPE:
74    (vi) ORIGINAL SOURCE:
75    (ix) FEATURE:
76
77    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
78
79  GAATTCATCA TTTTTCANGT CCTCAAGTGG ATGTTTCTCA TTTNCCATGA TTTTAAGTTT        60
80  TCTCGCCATA TTCCTGGTCC TACAGTGTGC ATTTCTCCAT TTTNCACGTT TTNCAGTGAT       120
81  TTCGTCATTT TCAAGTCCTC AAGTGGATGT TTCTCATTTN CCATGAATTT CAGTTTTCTN       180
82  GCCATATTCC ACGTCCTACA GNGGACATTT CTAAATTTNC CACCTTTTTC AGTTTTCCTC       240
83  GCCATATTTC ACGTCCTAAA ATGTGTATTT CTCGTTTNCC GTGATTTTCA GTTTTCTCGC       300
84  CAGATTCCAG GTCCTATAAT GTGCATTTCT CATTNNCAC GTTTTTCAGT GATTTCGTCA        360
85  TTTTTTCAAG TCGGCAAGTG GATGTTTCTC ATTTNCCATG ATTTNCAGTT TTCTTGNAAT       420
86  ATTCCATGTC CTACAATGAT CATTTTTAAT TTTCCACCTT TTCATTTTTC CACGCCATAT       480
87  TTCATGTCCT AAAGTGTATA TTTCTCCTTT TCCGCGATTT TCAGTTTTCT CGCCATATTC       540
88  CAGGTCCTAC AGTGTGCATT CCTCATTTTT CACCTTTTTC ACTGATTTCG TCATTTTTCA       600
89  AGTCGTCAAC TGGATCTTTC TAATTTTCCA TGATTTTCAG TTATCTTGTC ATATTCCATG       660
90  TCCTACAGTG GACATTTCTA AATTTTCCAA CTTTTTCAAT TTTTCTCGAC ATATTTGACG       720
91  TGCTAAAGTG TGTATTTCTT ATTTTCCGTG ATTTTCAGTT TTCTCGCCAT ATTCCAGGTC       780
92  CTAATAGTGT GCATTTCTCA TTTTTCACGT TTTTCAGTGA TTTCGTCATT TTTTCCAGTT       840
93  GTCAAGGGGA TGTTTCTCAT TTTCCATGAG TGTCAGTTTT CTTGCTATAT TCCATGTCCT       900
94  ACAGTGACAT TTCTAAATAT TATACCTTTT TCAGTTTTTC TCACCATATT TCACGTCCTA       960
95  AAGTATATAT TTCTCATTTT CCCTGATTTT CAGTTTCCTT GCCATATTCC AGGTCCTACA      1020
96  GTGTGCATTT CTCATTTTTC ACGTTTTTCA GTAATTTCTT CATTTTTTAA GCCCTCAAAT      1080
97  GGATGTTTCT CATTTTCCAT GATTTTCAGT TTTCTTGCCA TATACCATGT CCTACAGTGG      1140
98  ACATTTCTAA ATTATCCACC TTTTTCAGTT TTTCATCGGC ACATTTCACG TCCTAAAGTG      1200
99  TGTATTTCTA ATTTTCAGTG ATTTTCAGTT TTCTCGCCAT ATTCCAGGAC CTACAGTGTG      1260
```

INPUT SET: S36742.raw

```
100    CATTTCTCAT TTTTCACGTT TTTCAGTGAA TTC                              1293
101
102            (2)  INFORMATION FOR SEQ ID NO:2:
103
104        (i)  SEQUENCE CHARACTERISTICS:
105           (A)  LENGTH: 1044 base pairs
106           (B)  TYPE: nucleic acid
107           (C)  STRANDEDNESS: single
108           (D)  TOPOLOGY: linear
109
110        (ii)  MOLECULE TYPE: Genomic DNA
111        (iii)  HYPOTHETICAL: NO
112        (iv)  ANTI-SENSE: NO
113        (v)  FRAGMENT TYPE:
114        (vi)  ORIGINAL SOURCE:
115        (ix)  FEATURE:
116
117        (xi)  SEQUENCE DESCRIPTION: SEQ ID NO:2:
118
119    AGGCCTATGG TGAAAAAGGA AATATCTTCC CCTGAAAACT AGACAGAAGG ATTCTCAGAA      60
120    TCTTATTTGT GATGTGCGCC CCTCAACTAA CAGTGTTGAA GCTTTCTTTT GATAGAGCAG     120
121    TTTTGAAACA CTCTTTTTGT AAAATCTGCA AGAGGATATT TGGATAGCTT TGAGGATTTC     180
122    CGTTGGAAAC GGGATTGTCT TCATATAAAC CCTAGACAGA AGCATTCTCA GAAGCTTCAT     240
123    TGGGATGTTT CAGTTGAAGT CACAGTGTTG AACAGTCCCC TTTCATAGAG CAGGTTTGAA     300
124    ACACTCTTTT TTGTAGTATC TGGAAGTGGA CATTTGGAGC GATCTCAGGA CTGCGGTGAA     360
125    AAAGGAAATA TCTTCCAATA AAAGCTAGAT AGAGGCAATG TCAGAAACCT TTTTCATGAT     420
126    GTATCTACTC AGCTAACAGA GTTGAACCTT CCTTTGAGAG AGCAGTTTTG AAACACTCTT     480
127    TTTGTGGAAT CTGCAAGTGG ATATTTGTCT AGCTTTGAGG ATTTCGTTGG GAAACGGGAT     540
128    TACATATAAA AAGCAGACAG CAGCATTCCC AGAAACTTCT TTGTGATGTT TGCATTCAAG     600
129    TCACAGAGTT GAACATTCCC TTTCATAGAG CAGGTTTGAA ACACACTTTT TGATGTATCT     660
130    GGATGTGGAC ATTTGCAGCG CTTTCAGGCC TAAGGTGAAA AGGAAATATC TTCCCCTGAA     720
131    AACTAGACAG AAGCATTCTC AGAAACTTAT TTGTGATGTG CGCCCTCAAC TAACAGTGTT     780
132    GAAGCTTTCT TTTGATAGAG GCAGTTTTGA AACACTCTTT TGTGGAATCT GCAAGTGGAT     840
133    ATTTGTCTAG CTTTGAGGAT TTCTTTGGAA ACGGGATTAC ATATAAAAAG CAGACAGCAG     900
134    CATTCCCAGA ATCTTGTTTG TGATGTTTGC ATTCAAGTCA CAGAGTTGAA CATTCCCTTT     960
135    CAGAGAGCAG GTTTGAACAC TCTTTTTATA GTATCTGGAT GTGGACATTT GGAGCGCTTT    1020
136    CAGGGGGGAT CCTCTAGAAT TCCT                                         1044
137
138
139
140            (2)  INFORMATION FOR SEQ ID NO:3:
141
142        (i)  SEQUENCE CHARACTERISTICS:
143           (A)  LENGTH: 2492 base pairs
144           (B)  TYPE: nucleic acid
145           (C)  STRANDEDNESS: single
146           (D)  TOPOLOGY: linear
147
148        (ii)  MOLECULE TYPE: Genomic DNA
149        (iii)  HYPOTHETICAL: NO
150        (iv)  ANTI-SENSE: NO
151        (v)  FRAGMENT TYPE:
152        (vi)  ORIGINAL SOURCE:
```

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/502,664A

INPUT SET: S36742.raw

153	(ix) FEATURE:					
154						
155	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:					
156						
157	CTGCAGCTGG	GGGTCTCCAA	TCAGGCAGGG	GCCCCTTACT	ACTCAGATGG	GGTGGCCGAG	60
158	TAGGGGAAGG	GGGTGCAGGC	TGCATGAGTG	GACACAGCTG	TAGGACTACC	TGGGGGCTGT	120
159	GGATCTATGG	GGGTGGGGAG	AAGCCCAGTG	ACAGTGCCTA	GAAGAGACAA	GGTGGCCTGA	180
160	GAGGGTCTGA	GGAACATAGA	GCTGGCCATG	TTGGGGCCAG	GTCTCAAGCA	GGAAGTGAGG	240
161	AATGGGACAG	GCTTGAGGAT	ACTCTACTCA	GTAGCCAGGA	TAGCAAGGAG	GGCTTGGGGT	300
162	TGCTATCCTG	GGGTTCAACC	CCCCAGGTTG	AAGGCCCTGG	GGGAGATGGT	CCCAGGACAT	360
163	ATTACAATGG	ACACAGGAGG	TTGGGACACC	TGGAGTCACC	AAACAAAACC	ATGCCAAGAG	420
164	AGACCATGAG	TAGGGGTGTC	CAGTCCAGCC	CTCTGACTGA	GCTGCATTGT	TCAAATCCAA	480
165	AGGGCCCCTG	CTGCCACCTA	GTGGCTGATG	GCATCCACAT	GACCCTGGGC	CACACGCGTT	540
166	TAGGGTCTCT	GTGAAGACCA	AGATCCTTGT	TACATTGAAC	GACTCCTAAA	TGAGCAGAGA	600
167	TTTCCACCTA	TTCGAAACAA	TCACATAAAA	TCCATCCTGG	AAAAAGCCTG	GGGGATGGCA	660
168	CTAAGGCTAG	GGATAGGGTG	GGATGAAGAT	TATAGTTACA	GTAAGGGGTT	TAGGGTTAGG	720
169	GATCAACGTT	GGTTAGGAGT	TAGGGATACA	GTAGGGTACC	GGTAGGGTTA	GGGGTTAGGG	780
170	TTAGGGGTTA	GGGTTAGGGT	TAGGGTTAGG	GTTAGGGTTA	GGGGTTAGGG	GTTAGGGTTA	840
171	GGGTTAGGTT	TTGGGGTGGC	GTATTTTGGT	CTTATACGCT	GTGTTCCACT	GGCAATGAAA	900
172	AGAGTTCTTG	TTTTTCCTTC	AGCAATTTGT	CATTTTTAAA	AGAGTTTAGC	AATTCTAACA	960
173	GATATAGACC	AGCTGTGCTA	TCTCATTGTG	GTTTTCAATT	GTAACCACAT	TGTGGTTTCA	1020
174	ATGTGTTTAC	TTGCCATCTG	TAGATCTTCT	TTGCGTGAGG	TGTCTGTTCA	GATGTGTGTG	1080
175	CATTTCTTGN	NTTTNGGCTG	TTTAACTTAT	TGTTTAGTTT	TAATAATTTT	TTATATATTT	1140
176	GAAGACAAAT	CTTTCTCAGA	TGTGTATTTG	CAAATATTTC	TTCAATATGA	GGCTTGCTTT	1200
177	TGTCTCTAAC	AAGGTCTCTT	CAGAGATAAC	TTAAATATAA	GAAATCCACA	CTGTCACTTC	1260
178	TTTTGTGTAT	ATCTACCTTT	TGTGTCATTT	GTTAAAATTC	ATTACCAAAC	CCAAAGGCAG	1320
179	ATAGCTTTTC	TTCTATTGTT	TCTTCTAGAA	ATTTGTATAG	TTTTGCATTT	TTAGTGTAAG	1380
180	GATGATTTTG	AGTGATTATT	TGTGTAAGTT	GTAAAGTTTT	CGTCTATATC	CATATCATTT	1440
181	CTTATGGTTT	CCAATTAATC	GTTCCCTCAC	TATTTTTGGG	AAAGACACAG	GATAGTGGGC	1500
182	TTTGTTAGAG	TAGATAGGTA	GCTAGACATG	AACAGGAGGG	GGCCTCCTGG	AAAAGGGAAA	1560
183	GTCTGGGAAG	GCTCACCTGG	AGGACCACCA	AAAATTCACA	TATTAGTAGC	ATCTCTAGTG	1620
184	CTGGAGTGGA	TGGGCACTTG	TCAATTGTGG	GTAGGAGGGA	AAAGAGGTCC	TATGCAGAAA	1680
185	GAAACTCCCT	AGAACTCCTC	TGAAGATGCC	CCAATCATTC	ACTCTGCAAT	AAAAATGTCA	1740
186	GAATATTGCT	AGCTACATGC	TGATAAGGNN	AAAGGGGACA	TTCTTAAGTG	AAACCTGGCA	1800
187	CCATAAGTAC	AGATTAGGGC	AGAGAAGGAC	ATTCAAAAGA	GGCAGGCGCA	GTAGGTACAA	1860
188	ACGTGATCGC	TGTCAGTGTG	CCTGGGATGG	CGGGAAGGAG	GCTGGTGCCA	GAGTGGATTC	1920
189	GTATTGATCA	CCACACATAT	ACCTCAACCA	ACAGTGAGGA	GGTCCCACAA	GCCTAAGTGG	1980
190	GGCAAGTTGG	GGAGCTAAGG	CAGTAGCAGG	AAAACCAGAC	AAAGAAAACA	GGTGGAGACT	2040
191	TGAGACAGAG	GCAGGAATGT	GAAGAAATCC	AAAATAAAAT	TCCCTGCACA	GGACTCTTAG	2100
192	GCTGTTTAAT	GCATCGCTCA	GTCCCACTCC	TCCCTATTTT	TCTACAATAA	ACTCTTTACA	2160
193	CTGTGTTTCT	TTTCAATGAA	GTTATCTGCC	ATCTTTGTAT	TGCCTCTTGG	TGAAAATGTT	2220
194	TCTTCCAAGT	TAAACAAGAA	CTGGGACATC	AGCTCTCCCC	AGTAATAGCT	CCGTTTCAGT	2280
195	TTGAATTTAC	AGAACTGATG	GGCTTAATAA	CTGGCGCTCT	GACTTTAGTG	GTGCAGGAGG	2340
196	CCGTCACACC	GGGACCAAGA	GTGCCCTGCC	TAGTCCCCAT	CTGCCCGCAG	GTGGCGGCTG	2400
197	CCTCGACACT	GACAGCAATA	GGGTCCGGCA	GTGTCCCCAG	CTGCCAGCAG	GGGGCGTACG	2460
198	ACGACTACAC	TGTGAGCAAG	AGGGCCCTGC	AG			2492
199							
200	(2) INFORMATION FOR SEQ ID NO:4:					
201							
202	(i) SEQUENCE CHARACTERISTICS:					
203	(A) LENGTH: 28 base pairs					
204	(B) TYPE: nucleic acid					
205	(C) STRANDEDNESS: single					

INPUT SET: S36742.raw

```
206        (D) TOPOLOGY: linear
207
208    (ii) MOLECULE TYPE: Genomic DNA
209    (iii) HYPOTHETICAL: NO
210    (iv) ANTI-SENSE: NO
211    (v) FRAGMENT TYPE:
212    (vi) ORIGINAL SOURCE:
213    (ix) FEATURE:
214
215    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
216
217 GGGGAATTCA TTGGGATGTT TCAGTTGA                          28
218
219        (2) INFORMATION FOR SEQ ID NO:5:
220
221    (i) SEQUENCE CHARACTERISTICS:
222       (A) LENGTH: 29 base pairs
223       (B) TYPE: nucleic acid
224       (C) STRANDEDNESS: single
225       (D) TOPOLOGY: linear
226
227    (ii) MOLECULE TYPE: Genomic DNA
228    (iii) HYPOTHETICAL: NO
229    (iv) ANTI-SENSE: NO
230    (v) FRAGMENT TYPE:
231    (vi) ORIGINAL SOURCE:
232    (ix) FEATURE:
233
234    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
235
236 CGAAAGTCCC CCCTAGGAGA TCTTAAGGA                          29
237
238        (2) INFORMATION FOR SEQ ID NO:6:
239
240    (i) SEQUENCE CHARACTERISTICS:
241       (A) LENGTH: 47 base pairs
242       (B) TYPE: nucleic acid
243       (C) STRANDEDNESS: single
244       (D) TOPOLOGY: linear
245
246    (ii) MOLECULE TYPE: RNA
247    (iii) HYPOTHETICAL: NO
248    (iv) ANTI-SENSE: NO
249    (v) FRAGMENT TYPE:
250    (vi) ORIGINAL SOURCE:
251    (ix) FEATURE:
252
253    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
254
255 CCGCTTAATA CTCTGATGAG TCCGTGAGGA CGAAACGCTC TCGCACC     47
256
257
258
```

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION   *US/09/502,664A*

DATE: 02/05/2002
TIME: 03:23:23

INPUT SET: S36742.raw

Line   Error   Original Text